

Simpósio doutoral – Linguateca  
5 a 8 Maio 2005

Bolsa SAGI

Luís Cabral – Bolseiro Pólo Porto  
lcabral@letras.up.pt



# Sumário

- Sumário
- Objectivos da Bolsa
- Projectos
  - REPENTINO
  - Corpógrafo
  - SIEMÊS
  - SAGI

# Objectivos da bolsa

- Desenvolvimento de um projecto para gestão de interfaces Web
- Apoio ao nível de Engenharia de Software nos projectos em desenvolvimento no Pólo do Porto da Linguateca

# REPENTINO

REPositório para reconhecimento  
de ENTIdades NOmeadas



# REPENTINO

- *O REPENTINO é um sistema que armazena exemplos de entidades nomeadas, agrupando-as segundo uma taxonomia pré-definida.*
- *Criou-se a base de dados, interfaces Web de administração e para o utilizador, e scripts de processamento de estatísticas e visualização das mesmas.*
- *Cativação dos utilizadores através*
  - *Interfaces simples*
  - *Redireccionamento rápido para motores de pesquisa*
  - *Entidade do dia*
  - *Classificação por entidades submetidas*
  - *Exportação de dados em XML*

# Estatísticas

- Mais de 440.000 Exemplos de entidades
- 11 Categorias de Topo
- 101 Sub-Categorias

*Imagem aqui*

# SIEMÊS

*Sistema de Identificação de Entidades  
Mencionadas com Estratégia Siamesa*



# REPENTINO

- Baseado em duas estratégias distintas
  - Utilizando o REPENTINO
  - Utilizando regras de contexto
- Participação no HAREM

# Estratégias aplicadas

- REPENTINO

- Achou exemplo exacto? Optimo!
- Aplicação de métodos estatísticos aos exemplos semelhantes encontrados no REPENTINO

- Regras de contexto

- Valores numéricos
- Prefixos e sufixos de localizações, ordens de grandeza

- Déjà Vu

- Comparação com entidades já analisadas
- Busca de siglas

Corpógrafo



- Reestruturação do Corpógrafo
  - Estrutural
  - Visual

# SAGI

Sistema de Apoio à Geração de Interfaces

# Ferramentas para a Internet

- Código mais extenso
- Mais difícil de implementar
- Mais difícil de procurar erros
- Interfaces Simples
- Maior tempo de desenvolvimento nas interfaces
- Alterações mais complicadas

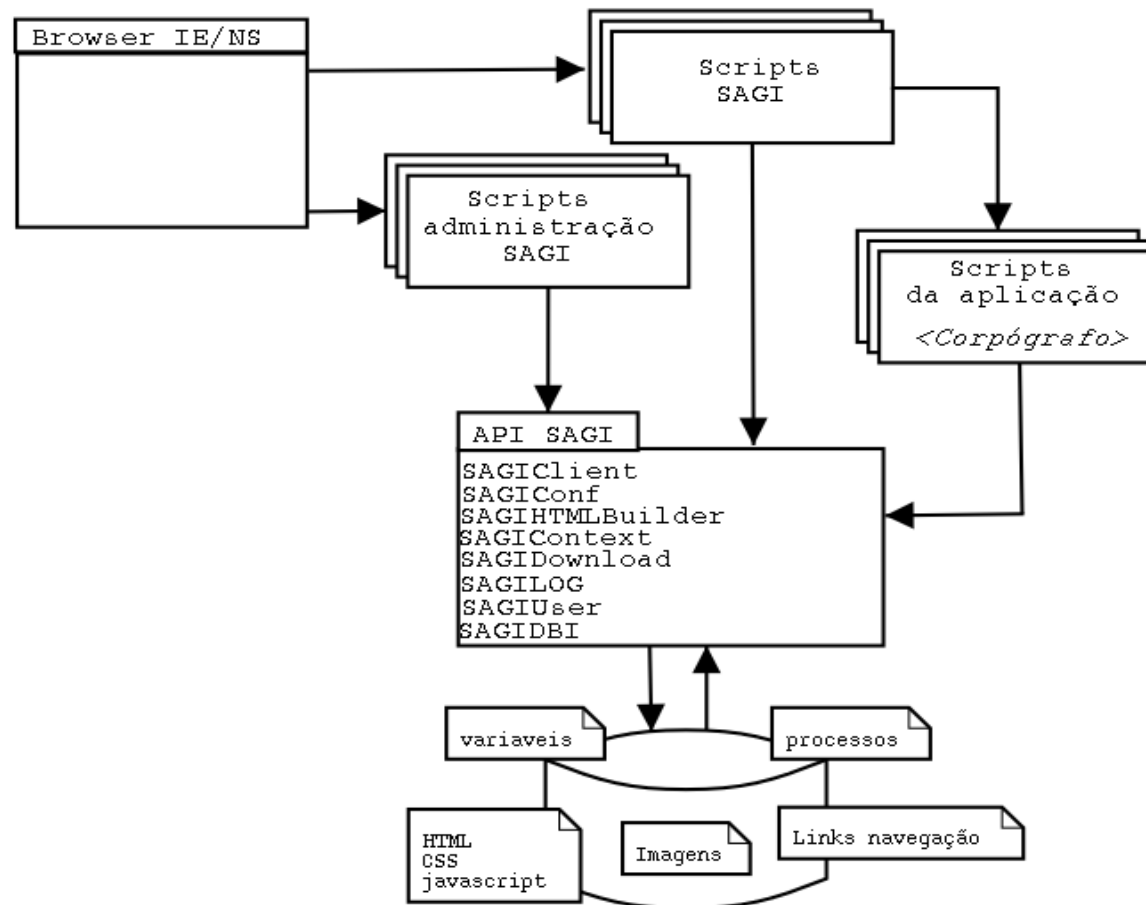
# O que é o SAGI?

- O SAGI é um sistema de Geração de interfaces Web padronizadas que gestão de conteúdos e processos
- Separação física entre a parte de interface e a parte lógica.
- Facilidade no desenvolvimento da aplicação
- Facilidade em proceder a alterações
- Melhoria da qualidade das interfaces
- Possibilidade de interfaces Multilanguages



- Tem a capacidade de:
  - Armazenamento de informação numa base de dados MySQL
  - Associar e organizar scripts independentemente da linguagem de desenvolvimento
  - Associa pedaços de código (Strips) aos scripts
  - Execução e Gestão de scripts
  - API em Perl para conectar as funcionalidades do SAGI

# Arquitetura do SAGI





# Administração

Administração do Corpógrafo - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://lab/cgi-bin/corpografo/user/admin/sagi/sagi.pl

Google CITESEER Support MySQL Perl Linguateca HTML temp e-mail Actual

Administração

- Strips Globais
- Gestão de Scripts**
- Interface
- SAGI Watch Dog
- Zona de testes
- Debug& Docs
- BLOB

Info Nó Info HTML Strips Info Ajudas

### Gestão de scripts

Nó actual : Ficheiros

Id do No: 30 Visível: ☒ Nivel 0

Nome: Ficheiros

No pai: 17 - Gestor

Descrição: Listagem dos ficheiros [v3]

URI/Caminho: /corpora/corpografo/web/scripts/user/gestorfich/ficheiros/

☐ URL ☒ System ☐ Fork ☐ Dump

Ícons do menu lateral:

Normal: GFILE

Seleccionado: GFILE

Salvar Apagar

☐ Avo adopta Netos

### Nó Corrente

Novo Nó

\* Voltar Gestor

### Links Adjacentes

* Formulário Adicionar Ficheiros Locais	N+ 31
* Editar Meta-Dados do Ficheiro	N 101
* Editar Texto	N 119
* Verificar Faseamento	N 120
* Formulário Download Ficheiro	N+ 128
* Operacao sobre Ficheiros	N 140
* Pop Up Ver Meta-Dados Ficheiro	N 151
* Adicionar Ficheiros[m]	N 255
* Dicionário Ficheiro	N 274
* processao upbad Blob	N 279

pc

### Links Proximos

* Ficheiros	V+ 30
* Corpora	V+ 70
* Escolhe Directorio	N 265
* Autores	V 269
* Organizações	V 270
* Multimédia	V+ 271

Done

Partly Cloudy, 19°C